REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
11/28/94

3. REPORT TYPE AND DATES COVERED
Final Report, 06/01/91–09/30/94

4. TITLE AND SUBTITLE
Optimal Universal Coding and Density Estimation

5. FUNDING NUMBERS
DAAL-03-91-G-0107

6. AUTHOR(S)
Yu, Bin

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Wisconsin
1210 W. Dayton Street
Department of Statistics
Madison, WI 53706

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARO 28722.17-MA

11. SUPPLEMENTARY NOTES
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Research progress has been made in the areas of empirical processes for mixing sequences, information theory, minimax estimation theory in source coding and non-parametric statistics, and Markov chain Monte Carlo (MCMC) methods.

Rates of convergence and Central Limit Theorems results have been obtained for empirical processes of dependent data, and they are very useful for studying statistical models with dependence structure. On the important MCMC convergence diagnostic problem, regeneration points have been introduced into the Markov chain using the split-chain technique; so has been a global approach based on the the estimated L^1 error and the Cusum path plot. Making connections between information theory and statistics, we obtained an information-theoretic result on the rate of convergence of a D-semifaithful code, and we also introduced non-parametric minimax lower bound techniques into bounding from below the redundacy in source codking.

14. SUBJECT TERMS
Density estimation, Empirical process, Markov chain Monte Carlo, Minimax, Minimum Description Length, Rate-distortion theory.

15. NUMBER OF PAGES
4

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Final Report on ARO Grant (DAAL03-91-G-0107)

awarded to University of Wisconsin-Madison

Principle Investigator: Bin Yu
Current Mailing Address:
Statistics Department
367 Evans Hall
University of California
Berkeley, CA 94720-3860

1 Statement of Problems Studied

Under ARO grant (DAAL03-91-G-007) for the period of October, 1990 through September, 1994, the principle investigator conducted research in the areas of empirical processes for mixing sequences, information theory, minimax estimation theory in source coding and non-parametric statistics, and Markov chain Monte Carlo (MCMC) methods.

2 Summary of Research Results

With partial support of the ARO grant, seven papers were written, some in collaboration with others, on empirical processes for mixing sequences and on Markov Chain Monte Carlo methods.

Statisticians have recently turned their attention to dependent data. One way to formulate dependence in data are the mixing conditions. Rates of convergence and Central Limit Theorems results for empirical processes of dependent data are very useful for studying statistical models with dependence structure. In [1], rates of convergence results are given for empirical processes of β-mixing sequences and using the techniques developed there, optimal rates of convergence results are obtained for density estimation error in the L^∞ norm and for mixing sequences in [5]. Jointly with M. Arcones at University of Utah, we provide in [2] a Central Limit Theorem of empirical processes for completely regular sequences under almost minimal conditions. In the same paper, limit theorems for U-processes are also given.

The Markov chain Monte Carlo (MCMC) methods are being studied intensively for both Bayesian and likelihood computations. The MCMC method enables us to obtain (dependent) samples from a target density from which direct sampling is difficult. Quantities of interest of the target distribution, such as mean, variance, and tail probabilities, can then be approximated using the MCMC sample. Since the target distribution is the stationary distribution of the constructed Markov chain, the success of the MCMC methods relies crucially on our ability to assess the convergence of the chain to its equilibrium. In the joint paper [10] with P. Mykland and L. Tierney of Univ of Chicago and Univ. of Minnesota respectively, we introduce regeneration points into the Markov chain using the split-chain technique, and therefore provide a way of diagnosing the convergence of the common MCMC schemes. In [11], a global approach for convergence diagnostic is introduced based on the estimated L^1 error of a kernel estimator by utilizing the information contained in the unnormalized target density form. In [12] (jointly with Mykland), a simple Cusum plot is used in the MCMC context to extract more diagnostic information from a single run of MCMC. Paper [5] addresses for the first time the density estimation problem in MCMC as well. It also gives a comparison between two sampling schemes in Markov Chain Monte Carlo simulation and an assessment of the MLE based on an approximate likelihood function using the Gibbs sampler.

Information theory and Statistics have always been closely related and are now finding more common grounds when researchers from both fields trying to communicate more with each other as evidenced by the joint IEEE-IMS workshop in Virginia last month. Five paper were written on information theory related topics, and in particular Minimum Description Length (MDL) principle were to statistical problems. The joint paper [6] with T. P. Speed at UC-Berkeley contains an information-theoretic result on the rate of convergence of a D-semifaithful code, extending part of existing results in the MDL literature from noisyless codes to rate-distortion theory (or D-semifaithful codes). In [3], a well-known hypercube technique due to Assouad is used to construct a minimax lower bound on the rate of redundancy of d-th order continuous Markov sources. The lower bound is explicit enough in d that the nonexistence of a redundancy rate in the continuous case can be deduced. Paper [9] makes connections between three very useful techniques in nonparametric minimax lower bound construction. In [8], we illustrate the application of the MDL principle to a typical stochastic learning problem, where the features range over a continuum. Moreover, we show that when the object we try to learn, e.g., the probability function of the weight, lies in a parametric class, the best rate at which we can estimate it (or in other words, the best rate at which we can "learn" about it) is the same as the complexity of the model, that is, minimum description length of the model. When the model class is much larger, say a smooth nonparametric class, the "learning" rate is much slower. Paper [7] concerns the problem of assessing the performance of model selection criteria in terms of two kinds of predictions in the context of normal linear regression. The particular selection criteria considered are AIC, BIC and Predictive MDL. It illustrates using a simple model that at the heart of of the problem of model selection is still the bias and variance trade-off, and no criterion is universally better than others.

3 Publication List

References

[1] B. Yu (1994). Rates of convergence for empirical processes of stationary mixing sequences. *Ann. Probab.* **22** 94-116.

[2] M. Arcones and B. Yu (1994). Limit theorems for empirical processes and U-processes for dependent data. *J. Theor. Probab.* **7** 47-71.

[3] B. Yu (1994). Lower bound on the expected redundancy for classes of continuous Markov sources. In *Statistical Decision Theory and Related Topics V*, S. S. Gupta and J. O. Berger (eds), 453-466.

[4] M. Arcones and B. Yu (1994). Limit theorems for Empirical processes under dependence. In *Proc. in Chaos expansions, multiple Itô–Wiener integrals and their applications.* 205-221.

[5] B. Yu (1993). Density estimation in the L^∞ norm for dependent data with applications to the Gibbs sampler. *Ann. Statist.* **21** 711-735.

[6] B. Yu and T. Speed (1993). A rate of convergence result for D-semifaithful coding. *IEEE Trans. on Information Theory* **39** 8813-820.

[7] T. Speed and B. Yu (1993). Stochastic complexity and model selection: normal regression. *J. Inst. Statist. Math.* **45** 35-54.

[8] J. Rissanen and B. Yu (1992). MDL learning. To appear in *Progress in Automation and Information Systems*, John Baras (ed), Springer Verlag.

[9] B. Yu (1994). Assouad, Fano, and Le Cam. To appear in *Festschrift in Honor of L. Le Cam on His 70th Birthday.*

[10] P. Mykland, L. Tierney, and B. Yu (1995). Regeneration in Markov Chain samplers. To appear in *J. Amer. Statist. Assoc.*

[11] B. Yu (1994). Estimating the L^1 error of kernel estimators for Markov samplers. Submitted to *J. Amer. Statist. Assoc.*

[12] B. Yu and P. Mykland (1994). Looking at Markov samplers through cusum path plots: a simple diagnostic idea. Submitted to *Biometrika.*

4 Scientific Personnel Involved

Bin Yu (Principle Investigator)

Ananda Sen (Research Assistant. Ph.D. 1993 under Professor G. Bhattacharyya. Bin Yu was on the Thesis Committee.)

Yonghong Yang (Research Assistant. Ph.D. Student)

Zhen Luo (Research Assistant. Ph.D. Student)